In the name of Allah, the Most Gracious, the Most Merciful

LA FACULTÉ

SEEK KNOWLEDGE

Copyright disclaimer

"La faculté" is a website that collects copyrights-free medical documents for non-lucratif use.
Some articles are subject to the author's copyrights.
Our team does not own copyrights for some content we publish.
"La faculté" team tries to get a permission to publish any content; however , we are not able to contact all the authors.
If you are the author or copyrights owner of any kind of content on our website, please contact us on: facadm16@gmail.com
All users must know that "La faculté" team cannot be responsible anyway of any violation of the authors' copyrights.
Any lucrative use without permission of the copyrights' owner may expose the user to legal follow-up.

Sur: www.la-faculte.net Espace E-learning pour apprentissage gratuit online Pour utilisation Non-lucrative

Partie 1 : Statistique descriptive

Généralités :

Introduction :

La stat. descriptive est une science qui consiste à récolter des données chiffrées sur des ensembles nombreux puis à analyser, commenter et critiquer ces résultats.

Vocabulaire :

Déf (1) : Population ou univers (pop.) : l'ensemble d'étude, notée Ω.

Déf (2) : Individu (ind.) : un élément de la pop.

Déf (3) : La taille de la pop. : le cardinal de la pop = le nbre d'ind. de la pop, noté $|\Omega|$.

$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$

$|\Omega| = n$.

Déf (4) : Échantillon (éch.) : sous-ensemble (partie) de la pop. Il est aléatoire si son prélèvement est soumis aux lois du hasard.

Déf (5) : Caractère : chaque ind. peut présenter plusieurs caractéristiques appellées caractères, notés X, Y, Z ...

ex : X : sexe → F, M } Modalités.
Y : âge.
Z : couleur des cheveux.

Déf (6) : Modalité : spécificité du caractère. ex : caractère X → F, M } Modalités.

Déf (7) : Nature : un caractère peut être :

- **Qualitatif** :
  - **Ordinal** : les mod. peuvent être classées (ordonnées) ex : intensité de la douleur.
  - **Nominal** : les mod. ne peuvent pas être classées ex : la couleur des yeux.
- **Quantitatif** : les mod. sont mesurables.
  - **Discret ou discontinu** : les mod. prennent des valeurs isolées ex : nbre d'enfants : 1, 2, 3, 0...
  - **Continu** : les mod. prennent des valeurs appartenant à une intervalle de $\mathbb{R}$. ex : poids.

facadm16@gmail.com Participez à "Q&R rapide" pour mieux préparer vos examens 2017/2018

Notes:

- Tout ce qui peut être mesuré avec un instrument de mesure est continu (ex: poids).

ε: erreur de calcul.

- Un car. quantitatif est aussi appellé variable statistique.

Chapitre 1: Statistiques simples:

On s'intéresse à l'étude d'un seul caractère.

I/ Définitions:

ex 1:

Une enquête effectuée dans un immeuble de 26 logts porte sur le nbre d'écoliers par logt. Les résultats de l'enquête sont les suivants:

0	1	0	5	2	3
0	1	1	1	0	1
2	1	2	3	1	2
3	1	1	3	1	4
2	2				

X = nbre d'écoliers / logt.

Mod. = {0, 1, 2, 3, 4, 5}.

k = nbre de modalités.

k = 6

n = effectif total.

Tableaux statistiques:

i	x_i	n_i	n_{ic}	f_i	f_{ic}	$n_i x_i$	$n_i x_i^2$
1	0	4	4	4/26	4/26	0	0
2	1	10	14	10/26	14/26	10	10
3	2	6	20	6/26	20/26	12	24
4	3	4	24	4/26	24/26	12	36
5	4	1	25	1/26	25/26	4	16
6	5	1	26	1/26	26/26 = 1	5	25
Total:		26		1		43	111

x_i = $i^{ème}$ modalité.

n_i = effectif de la $i^{ème}$ modalité

n_{ic} = effectif cumulé de la $i^{ème}$ modalité

$= n_1 + n_2 + n_3 + \cdots + n_i$.

ex: $n_{1c} = n_1$

$n_{2c} = n_1 + n_2$.

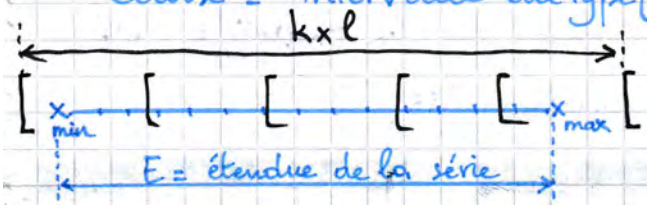
f_i = fréquence = n_i / n.

$f_{ic} = f_{1c} + f_{2c} + \cdots f_{ic} = \frac{n_{ic}}{n}$.

II/ Regroupement en classes:

On regroupe les données en classes de même amplitude.

Classe = intervalle du type $[a;b[$



$E = x_{max} - x_{min}$.

ℓ = amplitude (largeur de la classe)

k = nbre de classes.

Dans notre exemple: $E = 17{,}95 - 15{,}09 = 2{,}86$

$\ell = 0{,}5$.

Si $E = k \times \ell \Rightarrow k = \frac{E}{\ell} = \frac{2{,}86}{0{,}5} = 5{,}312$

on prend: $k = 6$.

$[15;15{,}50[$ $[15{,}50;16[$ $[16;16{,}5[$ $[16{,}5;17[$

$[17;17{,}5[$ $[17{,}5;18[$

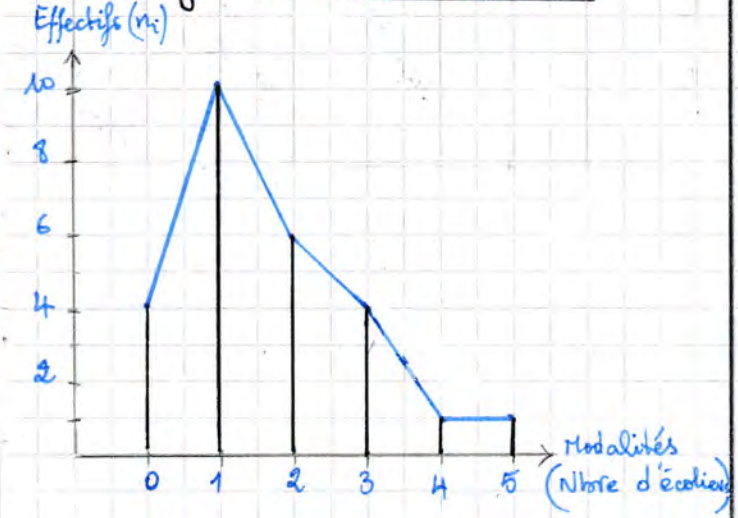
III/ Représentations graphiques:

A/ Caractères quantitatifs:

1- Caractère discret:

ex: les écoliers.

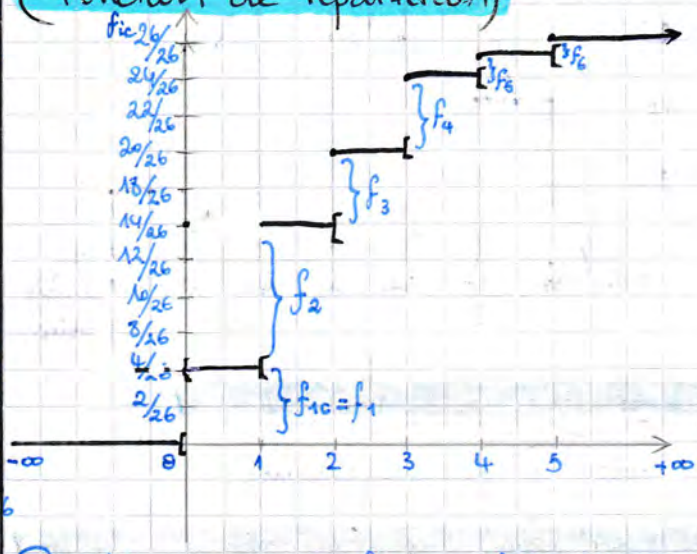
i/ Diagramme en bâtons:



ii/ Polygone des fréquences:

Voir le diagramme en bâtons.

iii/ Courbe cumulative des fréquences (Fonction de répartition)



Fonction en escaliers, elle est croissante continue à droite, discontinue à gauche.

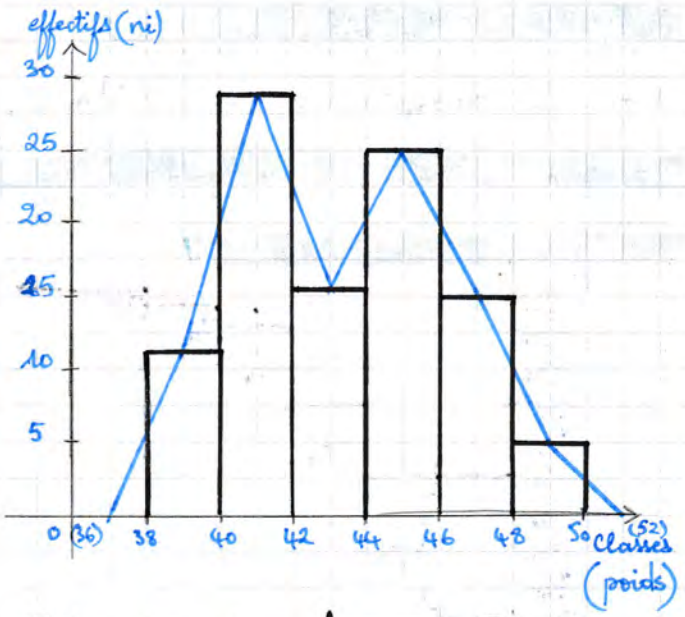
$\lim_{x \to +\infty} f(x) = 1$; $\lim_{x \to -\infty} f(x) = 0$.

2/ Caractère continu:

ex: le poids de 100 jeunes adolescents

	X				Y		
Classes	n_i	f_i	n_{ic}	Centre de classe	y_i	$n_i y_i$	$n_i y_i^2$
$[38;40[$	11	11/100	11	39	-2	-22	44
$[40;42[$	28	28/100	39	41	-1	-28	28
$[42;44[$	16	16/100	55	43	0	0	0
$[44;46[$	25	25/100	80	45	1	25	25
$[46;48[$	15	15/100	95	47	2	30	60
$[48;50]$	5	5/100	100	49	3	15	45
Total	100					20	202

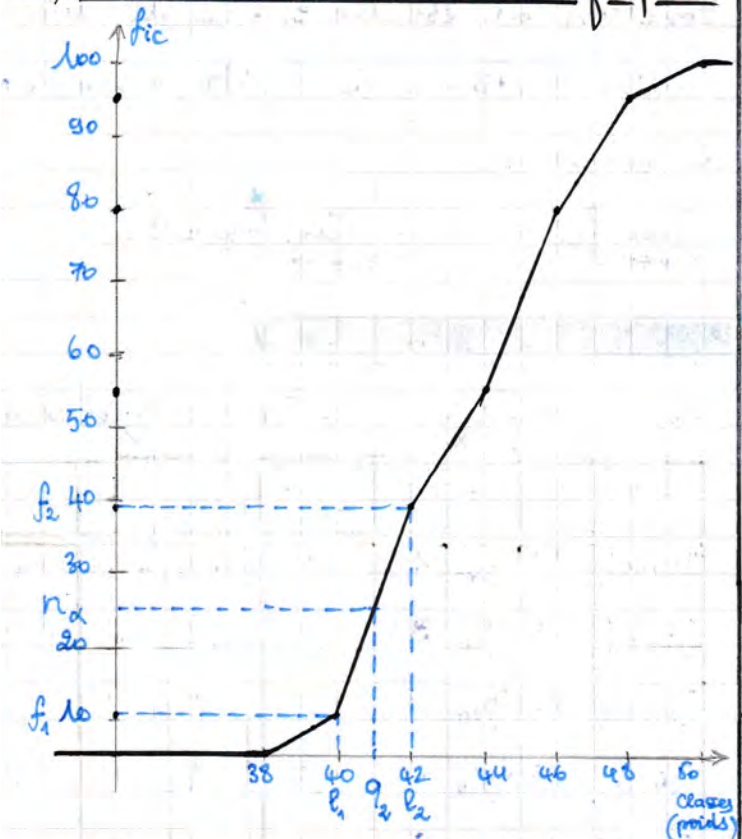
i/ Histogramme:



ii/ Polygone des fréquences:

voir l'histogramme.

iii/ Courbes cumulative des fréquences:



B/ Caractères qualitatifs:

ex: Répartition des principaux cancers chez les femmes dans la région d'Alger.

- Cancer du sein (30%)
- Cancer de l'appareil digestif (20%)
- Cancer des organes génitaux (20%)
- Cancer de la peau (6%)
- Cancer de la thyroïde (6%)
- Autres (18%).

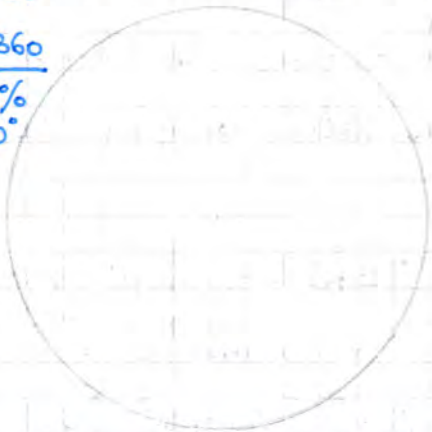
Diagramme en secteurs circulaires

100% → 360°.

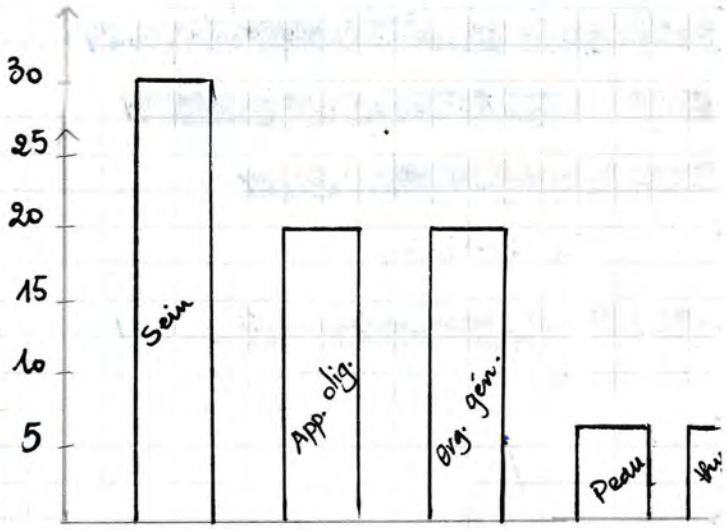
x% → θ.

$\theta = \frac{x\% \times 360}{100\%}$

$\theta = fic \cdot 360°$



Représentation en tuyaux d'orgue:



IV/ Paramètres de position et de dispersion:

A/ Paramètres de position:

1- Le mode: c'est la valeur la plus fréquente de la série statistique.

"Le mode est une modalité, non pas un effectif."

ex1: cas discret: les écoliers.

$M_0 = 1$. (eff. max = 10).

ex2: cas continu: le poids des ados

Le mode est le milieu (ou le centre) de la classe modale (celle qui a le plus grand effectif).

$M_0 = 41$. (eff. max = 28).

Une série statistique peut être uni-modale ou pluri-modale.

2- Les quantiles:

Soit la série statistique ordonnée

$X_1 \leqslant X_2 \leqslant \dots \leqslant X_n$

ex: les écoliers: 0000 1111111111 ...
$X_1 X_2 \dots$
3 4 5
$\dots X_{26}$

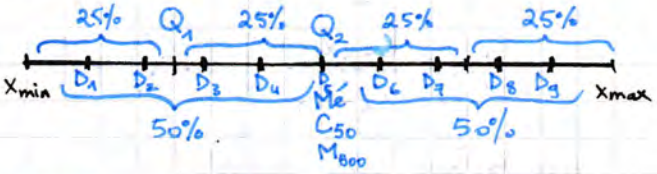
a/ Définition: Soit $\alpha \in]0;1[$. On appelle "quantile" d'ordre α de la variable statistique X, noté q_α la valeur numérique telle que le nbre d'observations inférieures à q_α ou égales soit égal à $n.\alpha$.

$n = 26$ $\alpha = 0,5$ $q_{0,5} = X_{14}$

$n\alpha = 13$

N.B.:

$q_{0,5}$ = la médiane = Mé = Q_2 (quartile) = $q_{0,5}$



$Q_1 = q_{0,25}$ $Q_3 = q_{0,75}$

- Il y a 3 quartiles.
- Il y a 9 déciles.
- Il y a 99 centiles.
- Il y a 999 milliles.

b/ Calcul pratique des quantiles:

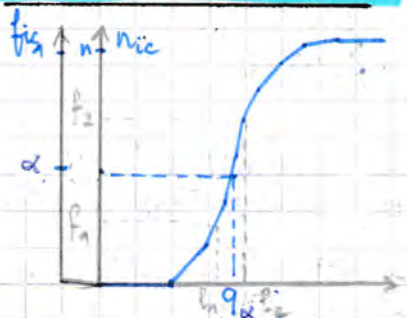
i/ caractère discret:

$$q_\alpha = \begin{cases} \dfrac{X_{n\alpha} + X_{n\alpha+1}}{2} & \text{si } n\alpha \in \mathbb{N} \\ X_{[n\alpha]+1} & \text{si } n\alpha \notin \mathbb{N} \end{cases}$$

$[n\alpha]$ = partie entière de $n\alpha$

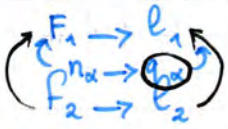
$[12,5] = 12$

ii/ Caractère continu:



Le quantile d'ordre α est l'abscisse du point qui a pour ordonnée $n\alpha$ ou α

$$q_\alpha = l_1 + \frac{l_2 - l_1}{f_2 - f_1}(n\alpha - f_1).$$



$$\frac{F_2 - F_1}{\ell_2 - \ell_1} = \frac{n_\alpha - F_1}{q_\alpha - \ell_1}$$

$$\Rightarrow q_\alpha - \ell_1 = \frac{\ell_2 - \ell_1}{F_2 - F_1}(n_\alpha - F_1)$$

$$q_\alpha = \frac{\ell_2 - \ell_1}{F_2 - F_1}(n_\alpha - F_1) + \ell_1$$

ex: les poids des ados.

$Q_1 = q_{1/4}$. $n_{1/4} = 100\left(\frac{1}{4}\right) = 25$

$11 < 25 < 39$

$40 < Q_1 < 42$

$$\frac{Q_1 - 40}{25 - 11} = \frac{42 - 40}{39 - 11}$$

$$\Rightarrow Q_1 - 40 = \frac{42 - 40}{39 - 11}(25 - 11).$$

$$Q_1 = 40 + \frac{42 - 40}{39 - 11}(25 - 11)$$

$Q_1 = 41.$

3- La moyenne arithmétique:

Elle est donnée par:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$$

k: nbre de modalités (cas discret)
↳ nbre de classes (cas continu).

x_i : → modalité (cas discret)
↳ centre de classe (cas continu)

n_i : effectif

n: effectif total

ex 1: cas discret (écoliers).

$\bar{X} = 1,65$ écolier / logt

ex 2: cas continu (le poid des ados)

$\bar{X} = 43,40$ kg.

B/ Paramètres de dispersion:

1- Variance et écart-type:

a/ Variance:

Elle est donnée par:

$$\text{Var } X = \frac{1}{n}\sum_{i=1}^{k} n_i (x_i - \bar{X})^2$$

Pour le calcul, on utilise:

$$\text{Var } X = \left(\frac{1}{n}\sum_{i=1}^{k} n_i (x_i^2)\right) - \bar{X}^2$$

b/ Écart-type:

Il est donné par:

$$\sigma_X = \sqrt{\text{Var } X}$$

* Remarque: $\text{Var } X \geq 0$

c/ Changement de variables:

Soit X une variable statistique, et a et b, 2 nbres réels ($\neq 0$).

Soit $Y = aX + b$, alors:

$$\begin{cases} \bar{Y} = a\bar{X} + b \\ \text{Var } Y = a^2 \text{Var } X \end{cases}$$

ex 1: les écoliers.

$$\text{Var } X = \left[\frac{1}{26}(111)\right] - 1,65^2 = 1,55$$

ex 2: le poids des ados.

$\text{Var } X = 7,92$

On pose: $X = aY + b.$

$$Y = \frac{X - b}{a}$$

b = centre de classe médian ou modal

$b = 43$

a = amplitude des classes

$a = 2$

$\text{Var } Y = \frac{202}{100} - (0,2)^2 = 1,98.$

$\text{Var } X = a^2 \text{Var } Y = 2^2(1,98) = 7,92$

2- Coefficient de variation :

On le note par CV, il s'écrit :

$$CV_x = \frac{\sigma_x}{\bar{X}}$$

ex 1 : les écoliers : $\bar{X} = 1,65$

$\sigma_x = \sqrt{\text{Var } X} = \sqrt{1,55} = 1,245$

$CV_1 = 0,75.$

ex 2 : les ados : $\bar{X} = 43,4.$

$\sigma_x = \sqrt{\text{Var } X} = \sqrt{7,92} = 2,81$

$CV_2 = 0,065.$

$CV_1 > CV_2$

$\Rightarrow$ Série 1 est plus dispersée que Série 2.

3- Intervalle inter-quartile :

$[Q_1 ; Q_3]$

C'est l'intervalle qui contient 50% des observations, qui se situent au centre de la série statistique, symétriquement par rapport à la médiane.

$(Q_3 - Q_1)$ = écart inter-quartile.